EM wave measurements of glottal structure dynamics

John F. Holzrichter*
Lawrence Livermore National Laboratory, Livermore, CA and
Dept. of Applied Science, UC Davis, Davis, CA

Lawrence C. Ng, Gerry J. Burke, Nathan J. Champagne II,
Jeffrey S. Kallman, & Robert M.Sharpe,
Lawrence Livermore National Laboratory
Livermore, CA

James B. Kobler and Robert E. Hillman
Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary
Boston, MA

John J. Rosowski
Easton Peabody Laboratory, Massachusetts Eye and Ear Infirmary
Boston, MA

*** Corresponding author: holzrichter1@llnl.gov**

Abstract:
Low power, radar-like EM wave sensors, operating in a homodyne interferometric mode, can be used to measure tissue motions in the human vocal tract during voiced speech. However, when used in the glottal region there remains uncertainty regarding the contributions to the sensor signal from vocal fold movements versus those from pressure induced trachea-wall movements. The signal source hypotheses were tested on a subject who had undergone tracheostomy 4 years ago as a consequence of laryngeal paresis, and who was able to phonate when her stoma was covered. Measurements of vocal fold and tracheal wall motions were made using an EM sensor, a laser-doppler velocimeter, and an electroglottograph. Simultaneous acoustic data came from a sub-glottal pressure sensor and a microphone at the lips. Extensive 2-D and 3-D numerical simulations of EM wave propagation into the neck were performed in order to estimate the amplitude and phase of the reflected EM waves from the 2 different sources. The simulations and experiments show that these sensors measure, depending upon location, both the opening and closing of the vocal folds and the movement of the tracheal walls. When placed over the larynx, the vocal folds are the dominant source. The understanding of the signal sources is important for many potential applications.

- **Person to whom inquires should be directed. Send to: L-1 , PO Box 808, Livermore, CA, 94550, or email to** holzrichter1@llnl.gov **.**

Introduction

Radar-like sensors transmitting electromagnetic (EM) waves have been used to measure properties of the human vocal system (Holzrichter 1995, 1998) during speech (see Fig. 1). These sensors

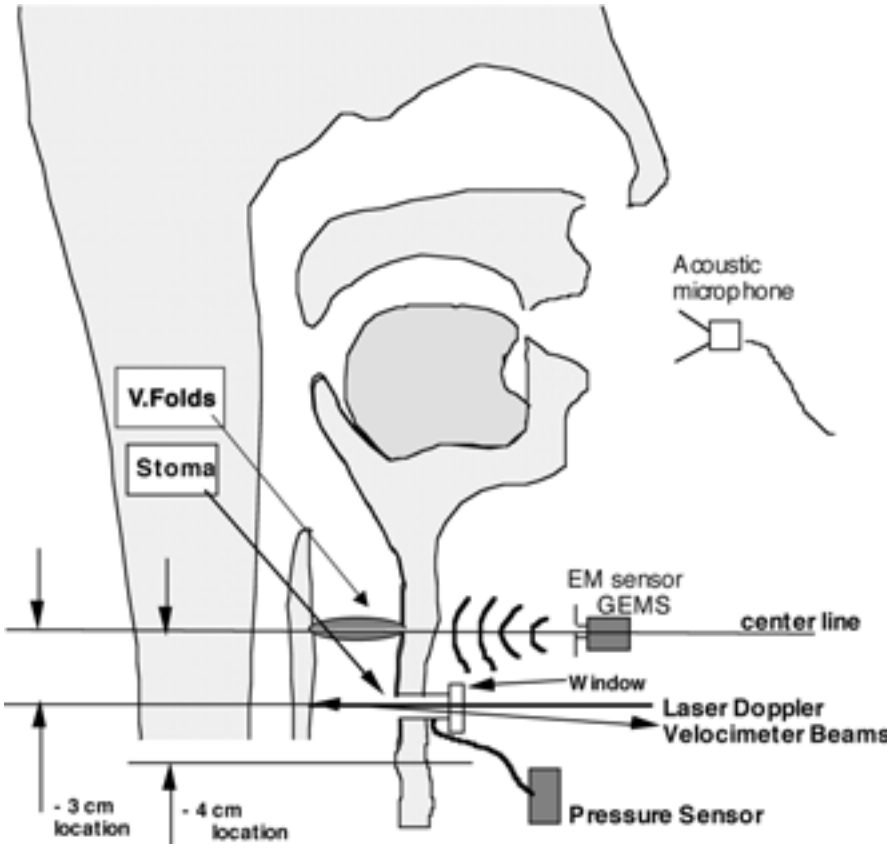

Fig. 1. Experimental layout showing position of the laser, the GEMs experimental location, pressure sensor, and acoustic microphone. EGG electrodes are not shown but are mounted in the normal location on the sides of the neck in the transverse plane of the vocal folds. Other locations for EM sensor placement are noted

work by transmitting very low power (<0.3 mW), EM waves, typically 2.3 GHz, into the neck or head toward the vocal articulators whose movements are being monitored. The absorption coefficient of 1-4 GHz EM waves in human tissue is 5 to 10cm^{-1}, which allows 10cm penetration of the waves into the body and reflection back to the sensor, with good signal to noise. The EM waves reflect from <u>all</u> dielectric and conductivity interfaces, that are associated with tissue-air or tissue-

tissue interfaces, and which are in the path of the propagating EM wave. The receiver circuitry detects the reflected EM waves using a homodyne (i.e., interferometric) technique (McEwan 1994, Skolnik 1990), then it averages, filters, and amplifies the signals such that they can be analyzed. Because voiced speech is associated with the movement of the vocal folds, at pitch frequencies nominally between 70 – 250 Hz, the signals from these structures and related pressure-induced tissue movements are easily distinguishable from stationary tissue interfaces by employing suitable band-pass filters. Typically the signal amplitude is directly proportional to a change in position of the targeted tissue interface, along the direction of EM wave propagation, and also to the reflecting area of the tissue interface. The particular sensor used for the experiments herein is called the GEMS sensor (Burnett 1999, McEwan 1996) because it is optimized for Glottal Electromagnetic Sensing. See below and Appendix A and B for additional details.

When the GEMs sensor is placed in front of the larynx as voiced speech is produced, the signals resemble vocal fold contact signals from electroglottograph (EGG) sensors, see Fig. 2 (Holzrichter 1998, Titze 2000). However EM waves propagate very differently than EGG generated currents, which flow longitudinally across contacting vocal folds (Rothenberg 1981, Titze 1990), and it has been shown that the EM signal does not depend on changes in vocal fold contact (Burnett 1999). One hypothesis is that EM waves reflect from the changing configurations of the air-tissue interfaces in the vocal fold structures (Holzrichter 1995). A second hypothesis is that the EM wave reflects from air pressure induced motions of the compliant posterior tracheal wall-air interface in the subglottal region (Burnett 1999). A third hypothesis is that refractive wave guiding by the vocal fold membrane, directs the wave across the fold contact surface similarly to the RF current from an EGG (Titze 2000).

Difficulty in pinpointing the source of GEMs signals occurs because the EM wave reflectivity is proportional to both the product of the targeted tissue's "transverse" area times its "longitudinal" movement along the path of the EM wave, and the source of reflection is ambiguous to 1/4 λ in distance. The targeted glottal-vocal-fold structure (as viewed front-on) is small in frontal area, about 20 mm^2, but it has an effective "movement" of approximately 5 mm during each open/close cycle of the glottis. In the case of a section of tracheal wall, the targeted tissue area A is relatively large (nominally 15 mm by 40

mm = 600 mm^2 in cross section), and the movement was estimated to be d = 0.1 mm (Burnett 1999, Svirsky 1997). Thus the product of (cross-section A) x (movement d) of the glottal or trachea wall tissues is about 60-100 mm^3 in both cases. If the above estimated values are correct, the reflection similarities would lead to GEMs signals with similar amplitudes for both types of tissue. However two strong tests of the tracheal wall hypothesis are that the directly measured wall movement must resemble the signals in Figs. 2 and 3, and the directly measured sub-glottal pressure vs. time pattern, Fig. 4, would need to be capable of moving wall tissue as expected. Similarly, for hypotheses 1 or 3 to be valid, analysis must show that the glottal opening in the vocal fold membrane must be capable of reflecting or refracting EM waves with sufficient amplitude to be measured by the GEMs sensor.

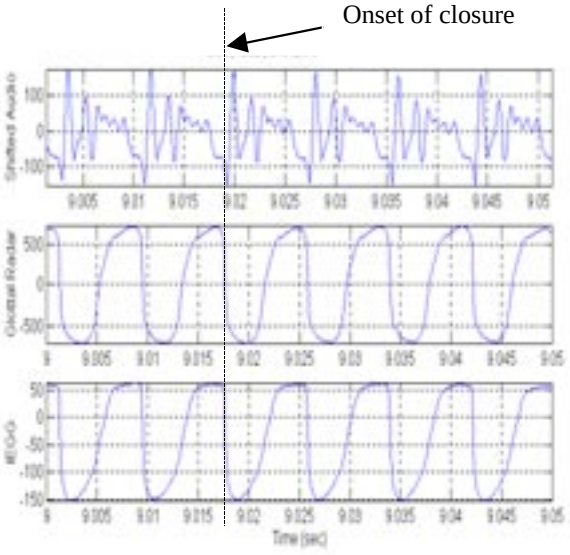

Fig. 2 Typical data showing simultaneous acoustic, GEMs sensor (i..e., "Glottal Radar trace), and Inverse EGG signals for a male speaker. Closure times are aligned.

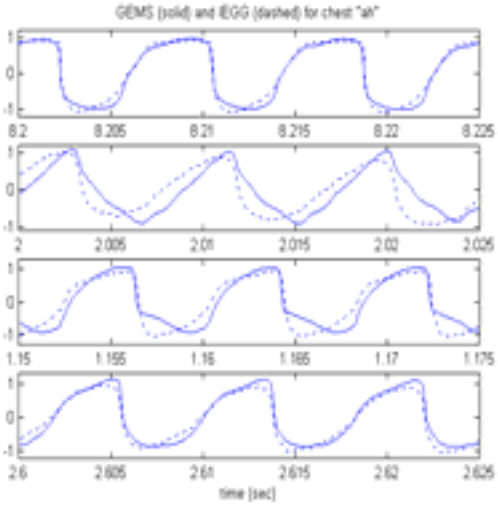

Fig. 3 Data from 4 separate male subjects comparing GEMs sensor data (solid) to EGG data (dashed). Closure times are not aligned. (with permission Titze 2000)

The primary purpose of this paper is to clarify the origins of the GEMs sensor signals by evaluating the hypotheses using the tests mentioned above. A laser-Doppler velocimeter was used, along with an EM sensor, to measure the posterior wall movements in the subglottal trachea of a subject who has a trachea-stoma located 3 cm below her laryngeal prominence (see Fig. 1). This female subject had undergone a tracheostomy, but she is now able to phonate when her stoma is covered (e.g., with a transparent plate for the measurements described here).

The pressure was measured by attaching a pressure transducer to the stoma valve. In addition, 2-dimensional (i.e., 2-D) and 3-dimensional (i.e., 3-D) numerical simulations were performed of EM waves propagating inward and then reflecting backward from both the posterior and anterior trachea walls, and from the vocal fold openings at several phases of the glottal cycle. The simulation data were related to corresponding GEMs signals to quantify the relative reflectivity of EM waves from the targeted tissue interfaces.

The data and interpretations in this paper provide a great deal of information on EM wave propagation and reflection from tissues internal to the neck, and by extension from other articulators in the vocal tract. This understanding is important because of the potential (Aliph 2002) of GEMs-like sensors to provide a wide range of timing information and estimates of voiced excitation functions during speech (Ng 1995).

II. Methods:

IIA. Experimental Configuration:

The subject for subglottal pressure and tracheal wall vibration measurements was a 57 year old female who had undergone a tracheostomy 4 years ago as a consequence of treatment for laryngeal paresis following thyroplasty with implantation of a silicone prosthesis. The paresis was partially resolved at the time of the recording, and the subject's voice, when occluding the stoma, was only slightly breathy. The experiments described below were approved by the human studies IRBs of the Massachusetts Eye and Ear infirmary and the Lawrence Livermore National Laboratory. A laser Doppler velocimeter (HLV-1000, Polytec PI, CA) was used to measure the movements of the subject's posterior tracheal wall by transmitting a laser beam through a transparent plastic cover, located over her stoma valve opening. Simultaneously the GEMs EM sensor was placed in one of two locations, depending upon the experiment. To test vocal-fold reflectivities, it was placed either on or slightly below the thyroid-prominence in the mid-sagittal plane. To measure tracheal wall movements, it was placed below the stoma opening, about 4 cm below the prominence, also in the mid-sagittal plane. The skin was marked to provide repeatable locations for sequences of measurements. In all experiments, the EM sensor was placed against the neck tissue to prevent skin motion. It was then moved in or out slightly to maximize the signal. In addition, data from an EGG sensor (Glottal Enterprises MC2-1), a subglottal pressure sensor, and an acoustic microphone placed in front of the lips were also recorded, as described below.

The subject habitually wore a size 6 tracheostomy tube with a Montgomery Tracheostomy Speaking Valve. For the experimental measurements, the speaking valve membrane was removed and a side opening in the valve body was drilled, to which a 6 cm long side tube, 3.0 mm ID, was attached. The tube was connected to an SMI 5552-015-D pressure transducer (Silicon Microstructures Inc., Fremont, CA). To enable phonation, the open end of the modified speaking valve was covered with a transparent plastic window through which the laser was aimed at the posterior wall of the trachea. The plastic window was heated by immersion in warm water prior to placements to minimize fogging. Most of the measurements were made with the subject reclining in a supine position.

The laser instrument has been used for many physiological structure measurements in the past, including making measurements through windows and through narrow openings (Rosowski 1999). Its internal velocity calibration is quite accurate and provides velocity output in volts, which using a given calibration of mm/second/volt, is converted to velocity in mm/sec in the data presented here. The velocity data was numerically integrated to obtain the change in position of the targeted tissue versus time. For example in a 1 ms time interval, a velocity of 5 mm/sec leads to a movement of 5 micrometers.

The signals from the laser interferometer, the EM sensor, the subglottal pressure transducer with connecting air tube, and the microphone were amplified and low-pass filtered to eliminate anti-aliasing (Axon Instruments Cyberamp 380) and were sampled at 20 kHz using a PC-based Axon Instruments Digidata 1200 data acquisition system. The pressure waveform was calibrated using a manometer and static pressures. The dynamic response of the pressure transducer and connecting plastic tube was measured using a closed system and an earphone driver, and found to be well behaved below 1 kHz. The microphone signal was calibrated in free field by reference to a calibrated Cooper-Rand electrolarynx sound source using a handheld RION NL-11 (Rion Inc.,Tokyo, Japan) sound level meter. The microphone was a Sony ECM 50 electret microphone. Calibration and analysis of signals was performed using Axoscope (Axon Instruments, Inc.) and Matlab (Mathworks, Inc.) software.

The GEMs sensor's response was re-calibrated using a mechanical vibration generator (Bruel and Kjaer, see ref.), and was compared to data obtained earlier by Burnett (1999). An aluminum metal target was vibrated sinusoidally, at frequencies from 100 to 200 Hz, and with amplitudes from 1 micrometer to 200 micrometers. These procedures reconfirmed the frequency response (> 1 kHz) and sensor maxima/minima sensitivity versus distance. They also enabled a comparison of signal amplitude levels from varying targets (e.g., a 1 cm by 2 cm plate shaped to simulate an area-segment of the tracheal wall) and they validated the linear proportionality of GEMs signal versus vibration amplitude at signal maxima. These procedures gave a GEMs signal calibration in air of about 5 mV/micron for a target vibrating about 4 cm from the sensor.

In the case of a vibrating air/tissue interface, buried deep inside a cylindrical dielectric body (e.g., the neck) an estimate of the reflection calibration is more difficult (see Appendix B). For a simpler geometry, such as tissue vibrations from a surface located about 1 cm inside the skin (equivalent to about 4-5 cm of air path) the calibration is about 2.5 mV/micron. This was obtained by measuring the inside surface of a subject's cheek with the laser velocimeter and comparing the movement to the corresponding GEMs signal. Another examples of a similar geometry is the anterior interface of the trachea. In the case of the posterior tracheal wall-air interface, which is hidden behind the trachea air tube, its vibrations are more difficult to detect. However, in the posterior wall calibration is calculated to be about 12x lower than that of the anterior wall, or 0.2 mV/micron. These calibrations are useful but probably only accurate to ± 50% . They are useful in estimating tissue vibration amplitudes from "buried" air-tissue interfaces which are not directly measurable with more direct techniques.

IIB. Results:

Measurements were taken on four separate occasions over the period of one and a half years. A very large amount of data was taken, with Fig. 4 illustrating one of the clearer and more complete sets of data obtained. In this experiment, five sensors were used while the subject phonated the phoneme /a/ as in "father". The data traces, from top to bottom in Fig. 4, are:

4A) The raw GEMs sensor output from its location on the laryngeal prominence.

4B) The velocity of the subject's posterior tracheal wall, measured 3cm below the laryngeal prominence. At the 3 ms time, the velocity becomes negative as the glottis opens and the pressure drops. At the time of vocal fold closure, at the 4 ms time, the negative wall velocity stops increasing (in the negative direction toward the laser) and begins to slow to zero. At the 5 ms time, the posterior wall begins to move away from the laser, i.e., it "balloons" and expands as the glottis closes and the subglottal air pressure rises. The nominal 0.1 ms duration velocity noise is due to short drop-out periods (i.e., scintillation) of the reflected signal back to the laser receiver.

4C) The numerical integration of the tracheal wall velocity trace 4B) gives the posterior tracheal wall position movement versus time. A positive value indicates that the posterior wall is "ballooning" in a posterior direction away from the laser sensor. The amplitude of the

movement is approximately 12 microns, which is much lower than earlier values used to estimate the GEMs signals. The time integrated data also show the "ballooning" movement lagging the velocity by 1/4 of the glottal time cycle. This wall movement shows very little correlation with the signal shown in trace 4A), and thus it can not be the source of the GEMs signal from the laryngeal location.

4D) The subglottal air pressure versus time shows 7 cm to 20 cm H_2O pressure variations versus time, with peak pressure excursions ranging from 5 cm to 15 cm H_2O. The average pressure change that moves the trachea walls is about 5 cm H_2O. The signal also illustrates distinct subglottal resonances (Ishizaka 1976, Cranen 1985, and Fredberg 1978). This pressure versus time data also shows almost no correlation with the GEMs data in trace 4A), indicating that the pressure is not moving a tissue surface that is in turn being measured by the GEMs sensor, at the laryngeal location. The data from trace 4C) indicates that the tracheal wall is responding slowly and out of phase to the rapid subglottal pressure signal resonances, with about a 1.5 – 2 ms time constant. This is expected from a membrane being driven above its natural resonances (about 30-40 Hz).

4E) The inverted EGG signal is especially useful for timing the onset of vocal fold closure and providing information on the quality of the vocal fold contact area versus time. The EGG signal was inverted to make it consistent with GEMS data from previous papers, e.g., Figs. 1 , 2, and 11. This IEGG data follows the GEMs data closely for these and previous experiments (Titze 2000), and was used to inspect data sets to identify those with normal vocal fold closure (e.g., Fig. 4). The subject often phonated with incomplete closure.

4F) The acoustic microphone signal shows the characteristic initial negative pressure signal as the folds close. The acoustic signal was advanced by 0.7 ms to correct for the slower traveling acoustic wave, from the vocal folds to the microphone located 10cm from the lips. The characteristic peaks of the acoustic signal are used to time align different data sets, as seen in Fig. 5.

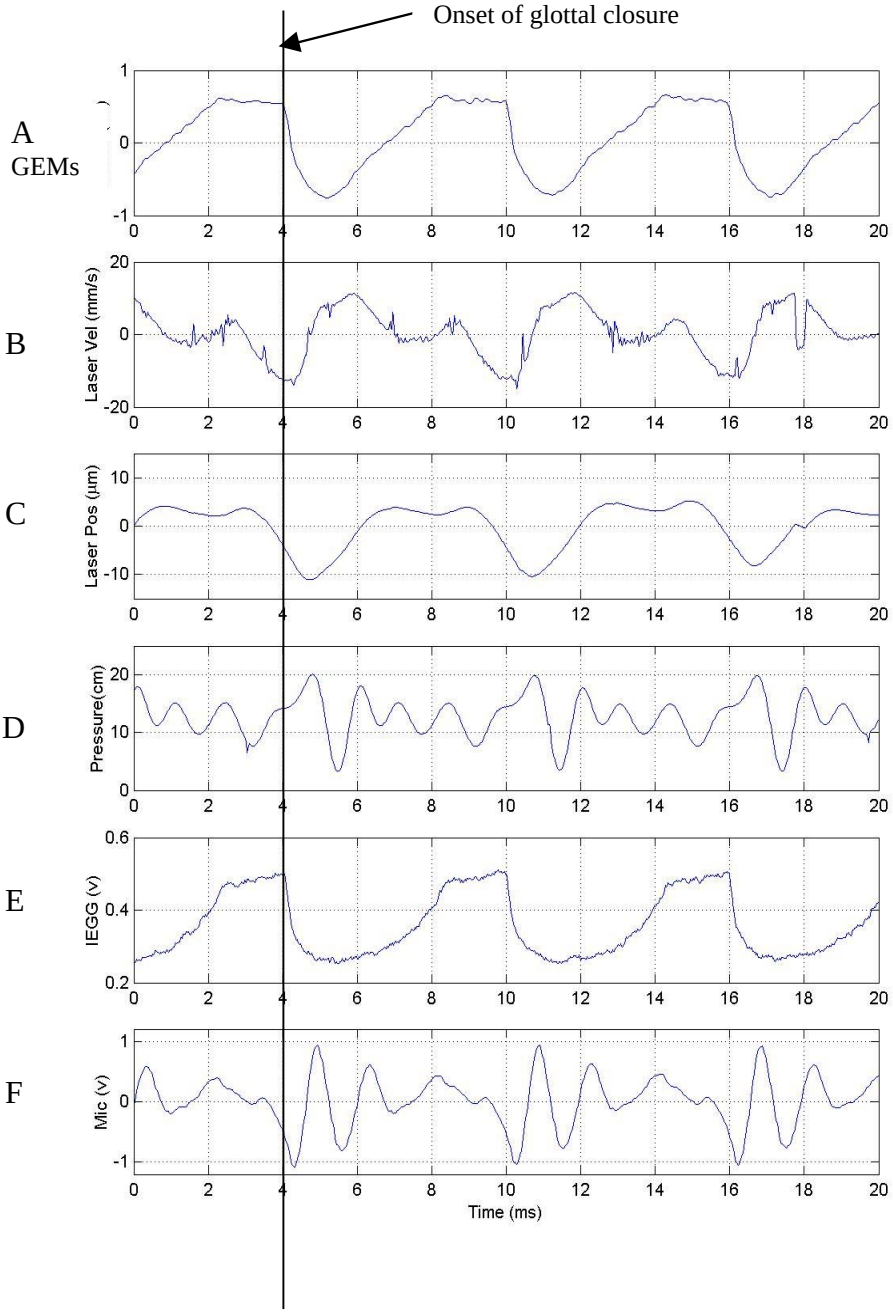

Fig. 4: Data set from subject with partial pariasis but phonating /ah/ very clearly. 6 traces are illustrated. The vertical line shows the time of onset of vocal fold closure; A) GEMs sensor with vertical direction indicating increased vocal structure reflectivity; B) Laser velocity signal from posterior tracheal wall, with vertical axis positive for tissue movement away from laser (i.e., posterior direction); C) Position versus time of posterior wall, with positive axis indicating "ballooning" as pressure increases; D) Subglottal pressure versus time in cm H_2O; E) inverted EGG signal; and F) Acoustic signal from microphone 10 cm in front of subject's mouth.

Another set of experiments (see Fig. 5) were performed to examine the differences between GEMs measured lower trachea wall movement and data from the laryngeal prominence location in Fig. 4. The lower trachea data, in the lower section of Fig. 5, were taken with the GEMs positioned 4 cm below the laryngeal prominence (about 1 cm below the stoma, see Fig. 1). Because of the spatial "congestion" of instruments within 1cm of the stoma, and due to some rf-interference, it was not possible to arrange for all of the instruments to be operative during those experiments in which the sub-stoma data was taken. The time alignment of the two data sets, from two separate sessions, was performed by correcting for small pitch variations, by aligning the acoustic-peak signatures. There are three important differences between the subglottal data shown in Fig. 5 and laryngeal prominence data of Fig. 4 and the upper part of Fig. 5. The first is that the sub-stoma GEMs signal is about 30 fold lower in amplitude, about 40 mV peak to peak in Fig. 5D, than the 1.2 V prominence signal in Fig. 5A. Using the GEMs calibration described above (2.5 mV/micron), the 40 mV signal indicates that the source is likely to be the anterior wall of the trachea, vibrating with a 10-20 micron amplitude. Secondly, the shape is consistent with a pressure-induced expansion and contraction, and shows no "sharp" cutoff of the type associated with vocal fold closure. Thirdly, the timing is consistent with the laser-measured posterior-wall data of Fig. 5B, except that it precedes it in time by about 1 ms. This timing offset may be due to the quite different types of trachea tissues being measured by the two instruments. Similar sub-glottal experiments have been conducted on male subjects at the Livermore Laboratory over the past years (in accord with earlier IRBs), which also show similar timing of the subglottal signal relative to the vocal fold closure signature. They also show similar relative signal amplitudes and show similar signal shape. In summary, the GEMs subglottal experimental data are completely consistent with EM wave reflections from the anterior tracheal wall "ballooning" due to changes in subglottal pressure, and as discussed later, are not consistent with simulations of EM wave reflections from the posterior wall.

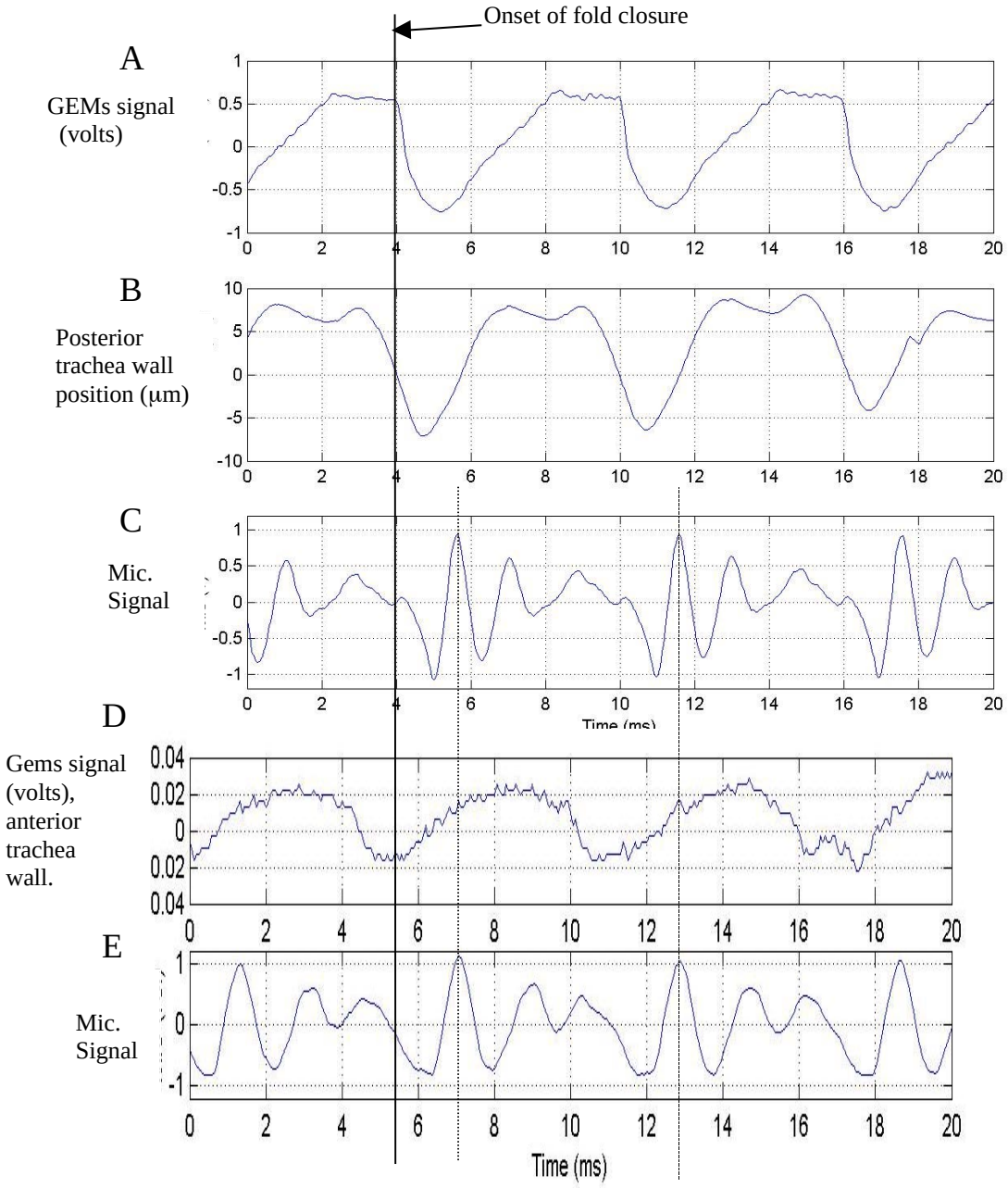

Fig. 5: Composite GEMs sensor data for the same speaker and same phoneme /ah/, but two different experiments. Upper data set is from Fig. 4A,C, D. Lower, subglottal set is the GEMs signal when located **4 cm below the prominence with the simultaneous acoustic data. The lower GEMs signal shape is consistent with a pressure driven tissue movement, just after the glottis closes. It shape is consistent with the posterior wall movement signal of 5B, but it rises and falls earlier. The two data files were time aligned using the acoustic signals from each file.**

III. Simulations

Two approaches were considered to simulate the reflections of EM waves from the posterior trachea wall and from vocal fold structures. First, a phantom, i.e., an analog simulator with a vibrating posterior wall and a vibrating vocal fold system, was considered. However, previous "pilot" mechanical mechanisms showed that many of a phantom's parts vibrated in response to the rapid 120 – 200 Hz motions of mechanical vocal structures. Very confusing signals were obtained from these vibrating parasitic interfaces. Instead, a numerical simulation approach was used based upon the availability of very advanced EM simulation codes. The relatively simple geometry of the tracheal tube and the glottal membrane with a circular glottal hole, allowed relatively simple zoning (Fig. 6) for a series of numerical experiments which are summarized below. The case of a membrane with a slot-shaped glottis required more complicated zoning, and is also discussed below.

IIIA. 2-D Simulations:

Two dimensional simulations were conducted to estimate the reflection of EM waves from the posterior wall of an empty tracheal tube. The simulation modeled the propagation of a 2.3 GHz EM plane wave through a 5 cm transverse aperture located just in front of an 8 cm radius, cylindrical dielectric neck ($\varepsilon = 28$) . The wave propagated through the aperture into the neck, and then traveled 1 cm further to the front surface of a 2.5 cm diameter tracheal tube. The wave then continued to the posterior wall of the trachea. Both tubes extended to infinity in the z-directions, to satisfy 2-D constraints. The 2-D simulations used a finite difference, EM code called TSARLITE (described by Kallman in the summary by Hawkins 1993). These simulations showed that the EM wave reaches the posterior wall by "evanescing" along the trachea air-tube walls. This is expected because the propagation of an EM wave in a high dielectric medium is energetically more favorable than in a low dielectric medium, such as the tracheal air tube. The 2-D simulations also show that about 10% of the radiated EM wave amplitude that enters the neck, through the simulated antenna aperture, is reflected back to the defining aperture from the anterior trachea wall, and 1% from the posterior wall. As these walls move, on the order of 10-20 microns due to air pressure changes, the modulation of the reflected waves is detected by the homodyne mixer in the GEMs sensor. The reflected amplitudes are

modulated by at about 2% and 0.8% respectively, due to both amplitude and phase changes in the reflected wave (see Figs. 8 & 9, and Appendix A and B).

The 2-D TSARLITE code could not be used for simulations of EM waves reflecting from the vocal fold structure, because the 0.4 cm thick membrane containing the vocal folds is limited in extent in the z direction, and can not be numerically simulated using 2-D approximations.

IIIB. 3-D simulations.

To follow the many physical processes as an EM wave propagates in a complex 3-D dielectric structure such as vocal folds in a tube, a computer code capable of 3-D simulations, such as the EIGER code (Sharpe 1997), was used. The Eiger code is based upon a very reliable "Method of Moments" approach (Miller 1992). This code was used to analyze the propagation of an EM wave internal to the neck tissue, whose layers of muscle and cartilage were approximated by an average dielectric constant, $\varepsilon_{average} = 25$, and zero conductivity (no loss in the tissues). Fig. 6 illustrates the shape of the modeled larynx region which was coded using 1500 nodes. A 2.3 GHz, EM plane-wave, y-polarized (in the transverse plane), with an amplitude of 75 V/M was launched from one side of the simulation (i.e., from within the dielectric coming from the right side in Fig. 6). The wavelength in the dielectric is $\lambda = 2.6$ cm. The EM wave propagated through the solid neck tissue to the front surface of the 1.5 cm diameter trachea (i.e., from the lower right of Fig. 6 into the "soup can" shaped model of the trachea section, which is 1.5 cm in diameter by 4 cm long). It then propagated around and, as well as being guided across the air tube by the high dielectric constant tissue of the vocal fold membrane. Fig 7 shows 4 examples of the EM wave being "scattered" across the membrane, and also back to the GEMs sensor. The numerical simulations used varying glottal openings in the membrane to determine their influence on the reflected EM wave. For purposes of the method of moments used for these simulations, the trachea air tube required "end caps" to keep the simulation volume finite in extent. Contributions to the E field at the point of measurement, from EM wavelets scattering from the caps, were tested to be sure that they did not contribute significant signals at the measuring point.

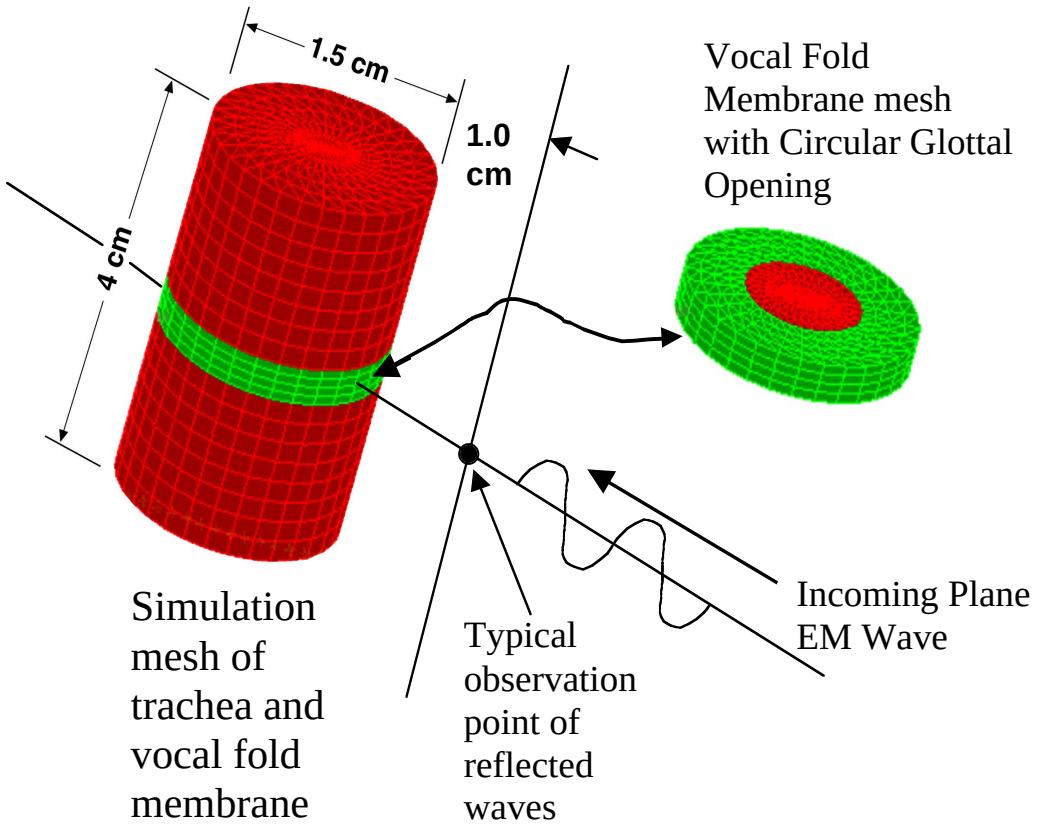

Fig. 6: Zoning of a 4 mm thick vocal fold membrane and a 1.5cm trachea tube for computer simulations of EM wave propagating into the vocal fold structures. The vocal fold membrane is simulated as a dielectric disk, ε = 25 , placed in the mid-transverse plane of the simulated "can shaped" air tube, which has end cap covers. The entire "can" structure is surrounded by a dielectric material with ε = 25, and with zero conductivity. The incoming EM wave is transversely polarized in the plane of the disk.

The Eiger code diagnostics enabled "measurement" of both amplitude and phase of the scattered part of the incoming EM wave at a desired observing point. This point was usually located 1cm in front of the anterior wall of the trachea in the mid-transverse plane (toward the incoming EM wave), a place in the experimental configuration where the EM sensor receiver antenna would commonly be located on the skin on the laryngeal prominence. Simulations of EM wave scattering to other locations, e.g., $\pm$1 and $\pm$2 cm up from the mid-plane along the neck surfaced plane, were examined to estimate the strength of the vocal fold scattering as the sensor was moved up and down the neck. In another set of experiments, the EM wave was brought into the vocal plane at an angle 20^0 to the plane (i.e., 70^0 from normal) to simulate oblique rays. This simulated non-normal placement of the GEMs sensor, where the skin surface is often canted by 10 - 20^0 due to the cartilage structure. These data indicate that the sensitivity of the EM sensor to vocal fold motion dropped as approximately a Gaussian curve of 1/2-width = 1.5 cm. Thus the EM sensor will detect EM wave reflections from the vocal folds several cm away from the mid-transverse plane.

The 3-D EIGER code was used primarily in the differential mode, where the input wave was held constant and changes in vocal fold configurations would cause phase and/or amplitude differences in the scattered EM waves, measured at a point 1 cm in front of the anterior tracheal wall. The configurations that were simulated include:

- **a) a 1.5cm diameter hollow trachea with a solid 4mm dielectric membrane in place (see Fig. 7D) was used to simulate a base-line reflectivity 1cm in front of the vocal fold membrane plane.**
- **b) the 1.5 cm hollow trachea , Fig. 7A.**
- **c) same as a) but with several circular glottal hole sizes whose volume is equaled by a corresponding thickness in the membrane to maintain constant membrane-tissue volume. See Figs.7B-C .**

- **d) same as a) but with two different parabolic slots in appropriately thickened membranes.**
- **e) the above with reflected Ey measurement points $\pm$ 2cm, $\pm$1cm, and 0 cm (on the sagittal line) to estimate the degree to which vocal fold motion causes EM reflections in directions upward and downward from the transverse plane**
- **f) the above with the incoming plane wave tilted 20^0 from the transverse plane, to simulate the EM sensor placed under the laryngeal prominence pointed upwards at the 20^0 angle.**
- **g) a 1.3 cm diameter, hollow, membrane free trachea, which is located 1 mm off-center to approximate the rear wall moving inward by 2 mm toward the incoming EM wave. (The amplitude and phase of the EM wave reflection from measured wall movements of 20 microns, are scaled relative to those associated with the 2 mm movement).**

The images shown in Fig. 7 illustrate contours of scattering EM waves. These contours are converted to plots of EM wave amplitude and phase versus distance along selected directions. Figs. 8 and 9 show the phase and amplitude along the path from the trachea center and through the measurement location 1 cm in front of the tube (see white line in Fig. 7B illustrating the path). Two glottal models were used to simulate the results reported here. The first consisted of a simple expanding and contracting circular hole in a 4mm thick membrane (ε_{Ave} = 25), starting with a solid membrane (i.e., ratio of (glottal area)/(trachea area) = area ratio AR= 0). Next a hole with relative area ratio AR=1/4 was simulated, and finally a hole with an area ratio AR= 1/2 . The condition AR=1 is the empty tube. The second, more complex model used a double-parabolic opening that resembled an expanding and contracting curved-edge slot in the membrane. In both the circular and parabolic cases, as the glottal opening was increased in the simulation, the glottal membrane was thickened to represent the volume of tissue that moves to the side as the glottal space opens. This approach conserved tissue molecules in the path of the EM wave.

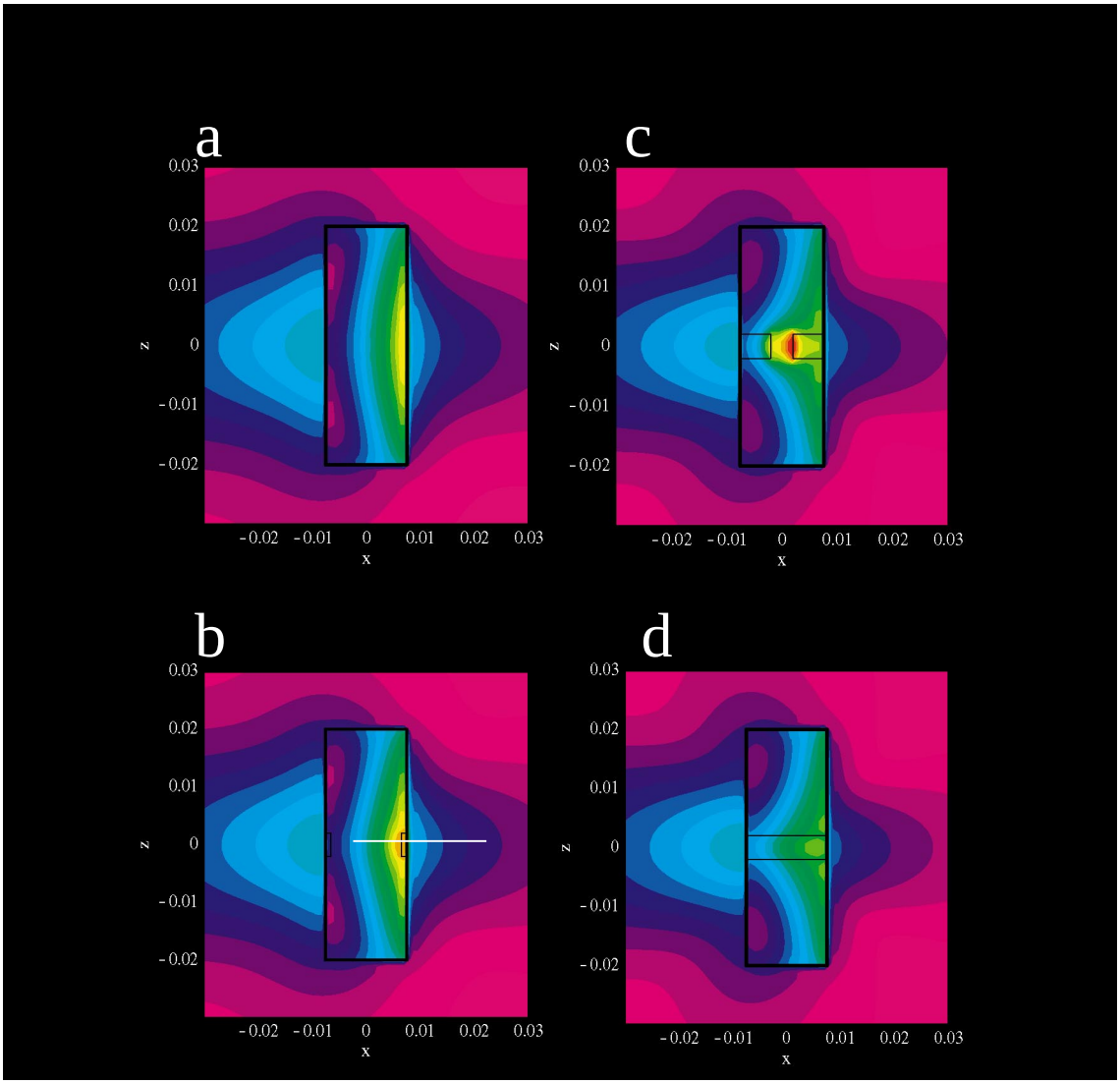

Fig. 7 : Four examples of simulations showing EM waves reflecting from 4 tracheal tube sections, each with a different glottal diameter in the central membrane. The outline of the vocal fold membrane is outlined in black along the transverse center line of the cylinder cross section. Four cases, a-d, are shown, with diameters of 1.5, 1.2 , 0.4, and 0 cm. Case d with the closed glottis best illustrates how the EM wave is carried across the tube by the 4 mm thick dielectric membrane, increasing its path length. In contrast, in the case 7a with completely retracted folds, where the 1.5 cm diameter glottis equals the tracheal wall diameter, the EM wave only weakly penetrates the anterior tube surface. In the intermediate cases, red and yellow colors indicate regions of high reflectivity. The white line in 7b shows the axis of "line-out" data shown in the next two figures.

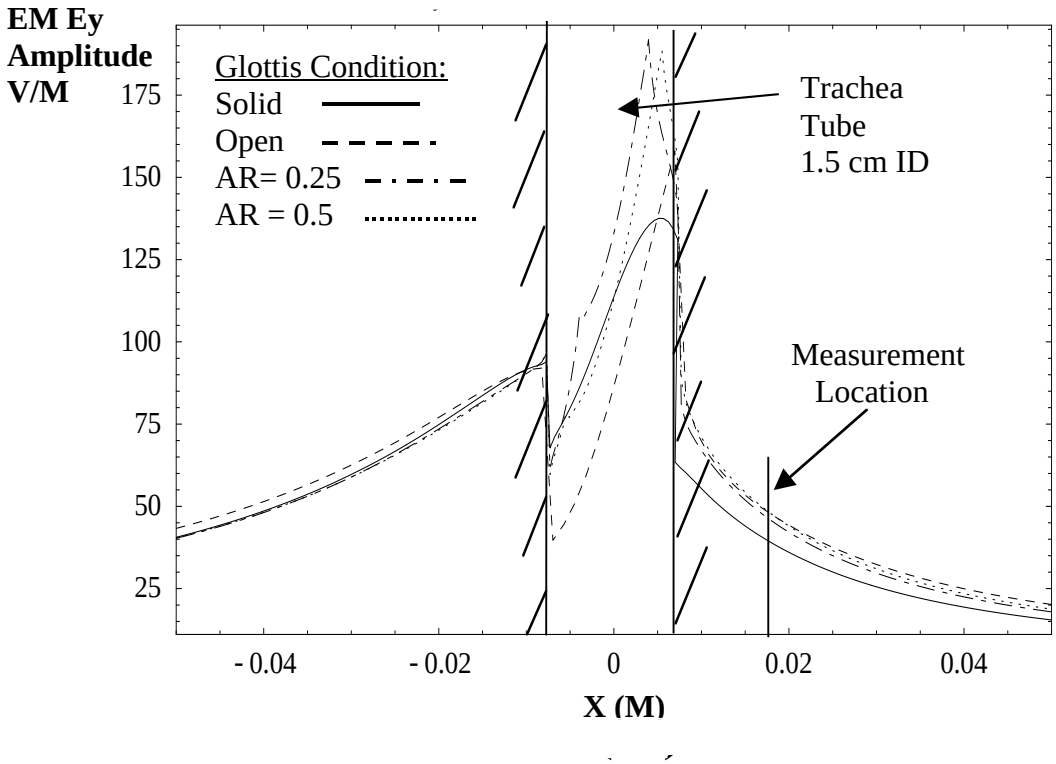

Fig. 8: Amplitude of the scattered EM wave both inside the trachea tube and also the reflected wave amplitude outside the tube, for 3 different circular glottal openings. This curve follows the scattered EM field along the mid-sagital center line of calculations. It gives a "line-out" along the white line (in Fig. 7b) showing the scattered EM wave amplitudes across the contour lines of data sets similar to those shown in the images displayed in Fig. 7. At the "measurement location" above, the intensity of the reflected EM wave depends upon the structure of the glottal membrane. The solid line is for a closed glottis in a 4mm thick membrane, the dot-dashed line is for a 0.75 cm hole (area ratio = 0.25), and the dotted line is for a hole diameter of 1.06 cm (area ratio = 0.5), dashed line is completely open .

The measurement location at 1.75cm from the trachea center is also 1 cm in front of the trachea's anterior wall. This is normally the location where the GEMs sensor is placed to transmit and receive the 2.3 GHz EM wave from the vocal fold region. At this point the graphs show the reflected field amplitude to range from 42 for closed folds, 45 with AR = 0.25, and 47 with AR=0.5, and 48 for an open tube trachea.

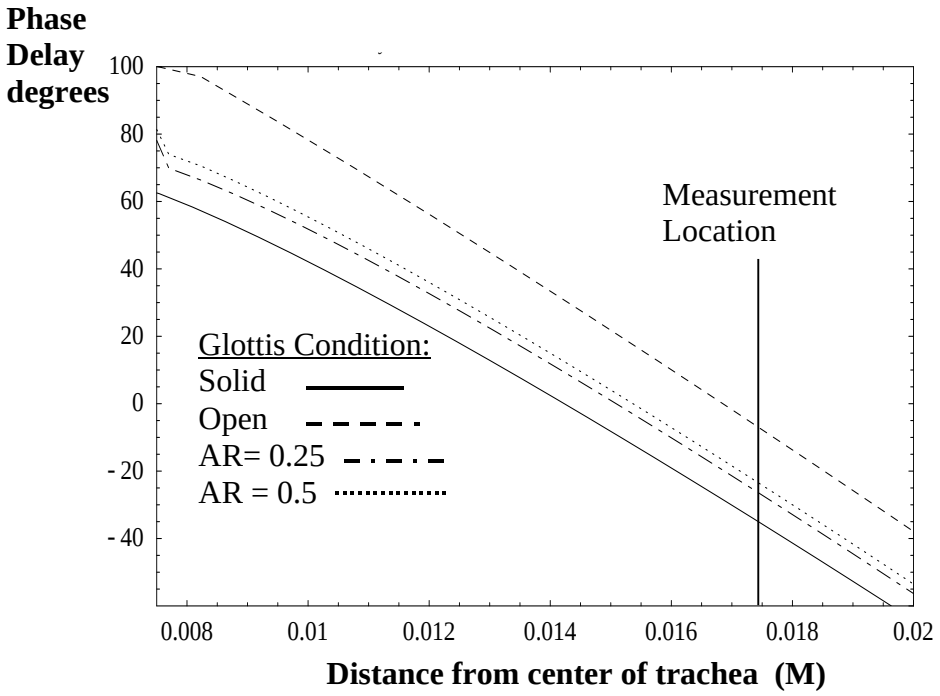

Fig. 9: The phase, in degrees, of the reflected EM wave versus distance and versus glottal condition from a measuring point to the trachea tube center. The left ordinate line is positioned at the anterior trachea wall. The phase increases (becomes more negative) as the glottis closes and "draws" the EM wave further inside the trachea tube. Conditions are constant input wave phase and amplitude, with reflected phase shown as functions of distance from trachea and condition of trachea.

The phase versus glottis conditions in Fig. 9 are the phase changes due to trachea scattering changes, at a fixed measuring point along the x axis. The phase changes at measurement location 1.75 cm (i.e., 0.0175 M or1 cm from the anterior tracheal wall) are the values used below for GEMs signal calculations. The convention in the Eiger code is that the phase is increasingly negative as the distance increases from the source to the measuring point. For example at the noted measurement location, the phase is increasingly negative as the glottis closes. This means that a fraction of the EM wave is carried deeper into the trachea tube as the glottis closes, thus increasing the path of the wave. The phase changes for the three glottal conditions shown are -36^0, -27^0, -25^0, and -8^0 for the empty tube. These significant phase changes are the primary source of the signal in the GEMs homodyne detection system,

where the output of the signal from the receiver detector is:

Equation (1): $S_{mixer} = const. * A * \cos \phi$

and A is the received signal intensity and ϕ is the phase referenced to a zero crossing (see Fig. 12) .

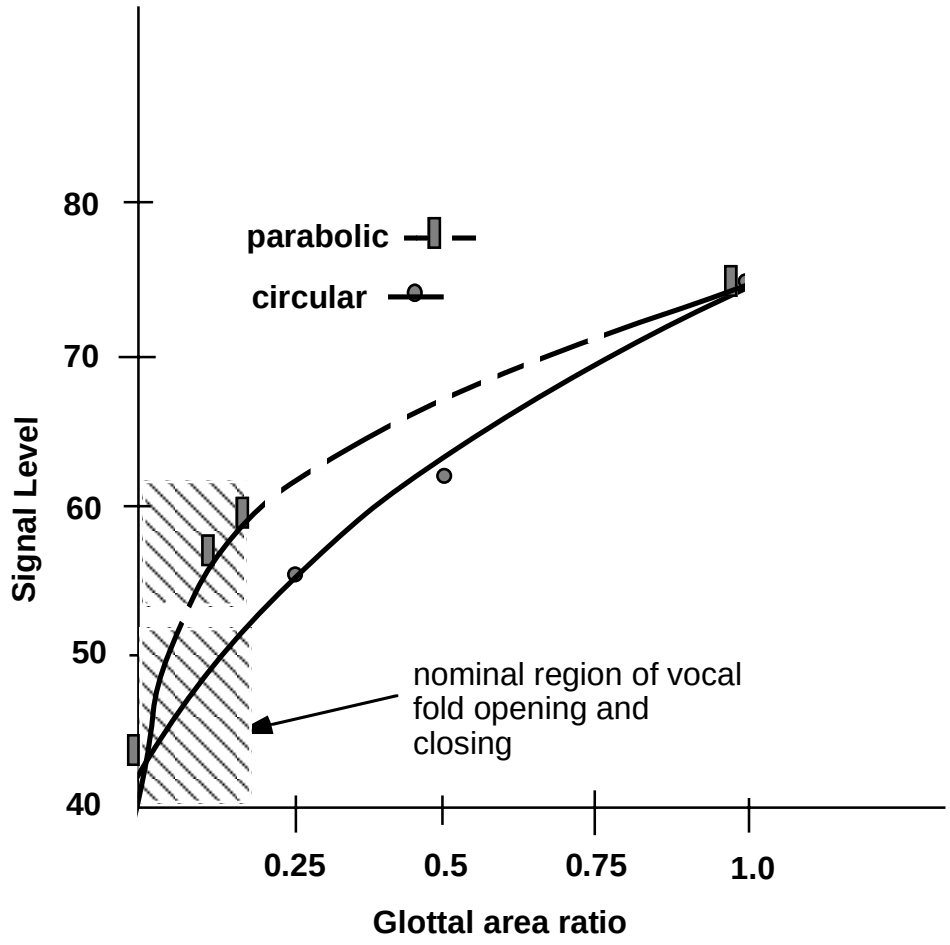

Fig. 10: Smoothed data from simulations showing the change in reflected EM wave intensity from circular glottal openings back to a receiver point located 1 cm in front of the anterior tracheal wall. The zero point, i.e., the reference signal, is the reflected signal strength of 42 V/M at the detector location, when the glottis is closed. The reflected signal change, from both circular and parabolic glottal openings, is a function of both amplitude and phase change from the reference signal. The ratio of glottal area to tracheal area (i.e. the Area Ratio) oscillates nominally between 0 and 0.2 during normal phonation.

It is interesting to note the process by which EM waves reflect, or rather "don't reflect" from the glottal opening, in contrast to their reflection from an open trachea tube. As the very high dielectric constant of the vocal fold membrane "draws" the EM wave into the tube (see Fig. 7D), scattering occurs further from the sensor location. This process lengthens the propagation path (and increases the reflected wave phase delay) and when the wave scatters inside the tube, less returns back to the sensor. In summary, the role of the vocal fold membrane and glottis is less important as reflectors of EM waves back to the sensor, but rather they serve as a "disperser" of the incoming EM wave, showing greater loss. These effects are illustrated in Figs. 8 & 9 which show that the solid membrane (solid line) is most effective in increasing the phase delay and reducing the amplitude of the backward reflected EM waves. When the vocal fold membrane is abducted (7A), the phase delay is reduced and the reflection to the GEMs increased. Fig. 10 shows that the vocal fold opening and closing cycle modulates the reflected EM wave phase by up to 30% (e.g., signal changes of (62 – 42) / 62 = .33 corresponding to area ratio changes from 0.0 to 0.20). This is a very strong modulation of the reflected wave. The strength of the glottal open-close GEMs signals are so very strong and so easily detected that they satisfy the requirement (i.e. the "test") for hypothesis one to be valid—the GEMS signal from the laryngeal prominence is from glottal opening and closing. However, details of EM wave scattering from small parabolic slots require additional work because the zoning and the field configurations in such slots are difficult to make realistic.

Fig. 10 shows plots of several simulated GEMS sensor signal levels for both circular and parabolic glottal openings versus the ratio of the glottal area to tracheal area. The estimates use the fact that at the maximum of the AC-coupled sensitivity point of a homodyne sensor (see Appendix A), the reflected signal, S , is proportional to the homodyne "mixer" output, S, which is proportional to the field amplitude at the antenna and the cos ϕ . See equation (1) above and in Appendix A, Fig. 12. This equation holds as long as the signal is changing rapidly enough in time to pass through the high pass filters in the GEMs sensor. The GEMS high pass 70 Hz filter, is satisfied because the glottis opens and closes at 100-200 Hz rates. The condition of maximum filter throughput (i.e., maximum first derivative) occurs when cos ϕ = (approx) zero (i.e., when ϕ is near π /2 , or an odd multiple n of π /2). Further, the analysis

for Fig. 10 assumes that the "initial condition" is when the vocal folds are closed, and when the GEMs sensor has been adjusted so that the round trip phase ϕ = (approx.) n * π /2 , and thus cos ϕ is assumed to be zero. As the vocal folds open, the reflected signal increases due both to the increase in reflected amplitude and the decrease in phase path. Fig 10 is plotted with the "initial condition" being the simulated steady state signal amplitude level (i.e., 42 V/M from Fig. 8), which because of the initial phase condition leads to a zero signal, S, output. Each of the simulated signal points in Fig. 10 are calculated, using the above homodyne signal equation, using both the amplitude and phase change as shown in Figs. 8 and 9. (Note that the units in Figs. 8 for the amplitude are in field strength units, V/M, which become volts on the antenna. Inside the sensor circuits, the antenna signal becomes eventually a 1 volt peak-to-peak GEMs A/C output signal, as in Fig. 4A, which appears as a + 0.5 V to –0.5 V AC signal. In the simulations summarized in Fig. 10, for the parabolic glottis open ratio of 0.2, the 1 volt GEMs signal is proportional to the simulated signal value of 62 – 42 (where the zero point value is 42) for a resulting AC signal value of 20. When the glottis closes, the area ratio = 0 and the simulated AC signal in Fig. 10 drops to 42, which is the zero-signal value, thus the AC signal becomes = 42-42 = 0 . In practice, the simulated signal values can be made proportional to the output sensor signal, e.g.,

Equation (2): $S_{glottal}$ (V) = (simulated value – 42) / 20 Volts

In this formula, the GEMs signal becomes 1 volt when the glottal area ratio is 0.2. Other normalization can be used, the one chosen above is to allow reference back to the simulations in Figs. 8 & 9. Notice that the simulated signal value is not a linear function of the area ratio, except near the origin.

V. Discussion of data and simulations

V-A: Tracheal Walls:

Two specific types of experiments and simulations were conducted. One focused on the movement of the anterior and posterior sub-glottal walls of the subject's trachea and the other on the vocal

folds. The posterior trachea wall is discussed first because of its early hypothesized role in generating the larynx signals as it moves under cyclic subglottal pressure oscillation. The degree to which this wall moves depends upon the compliance of the "soft" posterior smooth muscle tissue. This wall area is located just below the cricoid cartilage, where the trachea cartilage becomes C-shaped. It is adjacent to the anterior esophagus wall, in the opening of the "C" shaped cartilage. The direct measurement of this tissue with the laser velocimeter shows 12 micron movements. The wall behind the prominence itself is quite stiff because of the cricoid lamina, and is not expected to move very much under air pressure pulsation.

In contrast, the movement of the anterior trachea wall was not initially considered because of its expected stiffness. However, an examination of a human trachea from a cadaver showed that the stiff "C" shaped, cartilage structure only occupied about 30% of the anterior tracheal wall area. The remaining 70% of the wall tissue filling the space between the "C" rings was found to be similar to other vocal tract wall tissue, and could easily be caused to move by the oscillatory subglottal pressure. Thus its movement is examined more thoroughly below.

The numerical simulations above enable relative GEMs signal estimates, by comparing the simulated laryngeal prominence (i.e., glottal) signals to those of the anterior and posterior tracheal walls. The simulations show that most of the reflectivity of an EM wave is from the front surface of the trachea tube, and that changes in glottal configuration or in wall radius due to vibrations are "perturbations" on the total reflected amplitude and, primarily, on the phase of the reflected wave. Under these "perturbation" conditions, the homodyne signal is proportional to small changes in the phase versus time, $\Delta\phi(t)$. Considering first the signal from the laryngeal prominence for a small glottal area ratio change, e.g. from 0.0 to 0.20 , which causes a phase delay of about 14^0 (see Fig. 9) and also a small reflected amplitude change of about 10%, which can be neglected for this estimate. This enables the reflected wave calibration of the GEMs sensor to be:

Equation (3): $C_{glottal}$ **= 1 volt/14 deg phase delay**

Using Fig. 9 the anterior wall phase delay due to a 10 micron wall movements can be estimated by moving the measuring point by 10 microns and doubling the phase change (because both the transmitter and receiver are moved together). It is about 0.44^0/ 10 microns of movement (with no amplitude change). To estimate the anterior wall movement calibration, this simulated phase delay is multiplied by the GEMS calibration in (3) :

Equation(4): $C_{anterior} = (0.44^0/10\mu m)*(1V/14^0) = 0.030$ V/ 10 μm

The GEMs signal from Fig. 5 (below the prominence) is 35 mV, and using the simulation derived calibration of 30mV/10microns a movement of 12 microns is obtained. Separately, the GEMs signal was calibrated against a laser velocimeter for a similar geometry, and its calibration is 2.5 mV/ micron. Using the measured 30-40mV signal (see Fig. 5), and using the analog calibration, a movement of 14 microns is estimated. This close agreement is very good considering all of the variables, and it indicates that in the subglottal region it is highly likely that the GEMs sensor is measuring the anterior wall movement.

Further, the posterior wall induced changes in the GEMs signal that would occur if it were measuring 10 micron excursions can be similarly simulated (graphs for this case are not shown). A 10 micron motion of the posterior tissue leads to a simulated round trip phase change of 0.018^0. By using the GEMS calibration (3) of 1 V/14 deg., then the expected signal from the posterior wall is:

Equation (5)

$C_{posterior} = 0.018$ deg/10micron * 1V /14 deg. = 0.0013 V / 10 micron

Using the direct laser measurements in Fig. 4, of 13 microns of posterior wall motion, and this calibration, the GEMs signal would be about 1.7 mV. This signal is substantially smaller than the GEMs measured signal from the subglottal region, about 40 mV, and it indicates that the GEMs sensor signal is not from the posterior wall. However it is undoubtedly the case that this posterior movement contributes about 5% of the total GEMs sensor signal.

Additional estimates of the tracheal wall motions under experimentally measured sub-glottal pressure cycles were made using

simple calculations of the expansion of a trachea tube of "smooth muscle", under steady state conditions. A standard thick-walled pipe formula (e.g., Roark 1975) was used with an inner radius a of 0.75 cm, and an approximate outer radius b of 2.5 cm.

Then the change in the inner radius due to pressurization is:

Equation (6): $\Delta a = (\Delta \text{pressure}) * (a / E) * [(b^2 + a^2)/(b^2-a^2) + \upsilon]$

The radius a is 0.75cm, the Δ pressure is 5 cm H_2O, Young's modulus E for smooth muscle is estimated to be about 190 psi (Funt 1990, Whirley 2001), and $\upsilon = 0.2$. This calculation leads to a cyclic wall deflection of about 5 microns, due to air pressure excursions of about 5 cm H_2O. However, it is possible that the posterior wall tissue is more compliant, perhaps by a factor of 2-3 than "standard" smooth muscle leading perhaps to 10-15 micron movements. A similar calculation, modeling the posterior trachea wall as a compressible plate, using a unit strain formula of $\varepsilon = \sigma / E$, yields a movement of approximately 10 micrometers per 2 cm thick wall for a 5 cm H_2O pressure. These estimates as well as tissue compliance (i.e., "K" values) estimated from our posterior tracheal wall measurements (i.e., 13 microns per 5 cm H_2O pressure change) give values of about $30x10^5$ dynes/cm. This is substantially stiffer than exterior side-neck tissue compliance data from Ishizaka, $5x10^5$ dynes/cm (1975). This Ishizaka neck data is also about 10-fold stiffer than tongue tissue data presented by Svirsky, $0.5x10^5$ dynes/cm (1997), upon which early tracheal wall motion was estimated. In summary, these static stiffness estimates, the 2-D and 3-D simulations, Ishizaka's data, and the experiments in this paper show that the posterior tracheal wall movement is approximately 5-15 microns, which is not sufficient to generate the GEMs signals at laryngeal prominence location.

Finally, a subsidiary laser velocimeter and GEMS experiment was performed to measure the interior cheek movement, inside the oral cavity, as one of the authors phonated the sound /m/ . The laser velocimeter beam was passed through a window mounted on a tube, that was held between the subject's lips, to measure the inner cheek wall motion. The inner cheek surface is positioned about 1 cm from the

outer skin surface, where the GEMs sensor was positioned. The intra-oral pressure was measured with the same transducer described above, and showed a pressure excursion of 5 cm H_2O. Both the GEMS and laser sensors measured inner cheek surface movements of about 10 and 20 micrometers respectively, similar to that expected from Ishizaka's (1975) cheek data. This measurement of intra oral cheek tissue provides more evidence that internal vocal tract tissue surfaces, under 5 cm H_2O pressure cycles, move cyclically with an amplitude of approximately 10-20 microns. This comparison with the laser, also enabled a GEMs calibration for planar tissue.

<u>V-B: Laryngeal Prominence-vocal folds:</u>

Experiments on the female subject described earlier in this study (and with previous subjects before this study under separate IRBs at the Lawrence Livermore Laboratory) show that the GEMs sensor, when located on the laryngeal prominence consistent with vocal fold opening and closing, and of magnitude 1 Volt. The signal shape is consistent with EM wave reflections caused by a hole or slot slowly opening and rapidly closing in the glottal membrane. For a parabolic glottis, with area ratios smaller than 10% (see Fig. 10), it appears as if it may be possible to describe the relationship between EM wave reflection and glottal area by a linear approximation. A small area-ratio regime occurs when the folds finalize their rapid closure, AR < 5%, generating the higher harmonics in the air-flow that are the source-function for voiced speech. Additional work is required to understand GEMs signals from this phase of fold closure because of two effects. First, it is known that EM wave reflectivity, from disk-like structures that are significantly smaller than the wavelength, drops off very rapidly, as the ratio of size/λ < 1 . The formula (Born and Wolf 1965) show a reduction in scattering amplitude that is a high power, n = 3-6 , of the structure-to-wavelength ratio, (structure-size / λ_{eff}) ^{n} , depending on the disk shape, attitude to the wave, and material. Thus the signal from the rapid closure of the vocal folds is likely being enhanced by this super-linear reduction in the reflectivity. Secondly, the effect of tissue conductivity on the reflectivity of the EM wave from a complex structure must also be considered. Conductive structures that contact (e.g., vocal folds) can change the field configuration of EM waves, and thus change the reflectivity, although nodule experiments (see below) do not indicate this effect.

The signal vs. glottal area data relate to early suggestions (Ng 1995) that the EM sensor signal can be associated with a voiced speech excitation function. This conjecture has proven to be very useful in signal processing work (Burnett 1999, Gable 2000, Ng 2000); however, more work is required to justify the validity and range of this assertion. Because of the expense of the needed simulations of EM wave scattering, from mm sized, slot-like, partially conducting structures, as the folds open and close and as internal conducting structures rearrange themselves, they will be accomplished at a later date.

The EM wave guiding effect of the membrane, as illustrated in the above simulations (see Fig. 7), explains some remaining problems and observations from previous work (Burnett 1999, Titze 2000). First, it shows how hypothesis three (presented in the introduction to this paper), referring to refraction of a slightly off-axis, in-coming EM wave, back to the GEMs sensor, becomes more like hypothesis 1. The simulations here show that the vocal fold membrane does not have enough of a "lensing" effect to refract the EM wave back toward the source. However off-axis and refraction effects are important and are accounted for in the simulations.

Another related observation concerns Burnett's (1999) investigation of a subject who had a nodule on one of the vocal folds. He noted that the nodule caused almost no change in the GEMs signal compared to a normal individual's signal. The models above indicate that such a nodule, attached to one of the vocal folds and exhibiting transverse motion across the forward propagating EM wave, will have a very small effect on the reflectivity of the EM wave. Also, he noted that when the nodule first contacted the opposing fold, no signal-signature of the event took place, and shortly thereafter, complete fold closure occurred accompanied by a typical GEMs closure signal. This observation speaks against vocal fold contact-conductivity playing a significant role in the GEMs signal. Another potential signal source is the movement of the entire glottal structure up and down by a few millimeters as sub-glottal pressure increases and decreases. This is also a transverse tissue motion, relative to the EM wave propagation, and is not expected to contribute significantly to the GEMs signal.

The guiding effect of the vocal fold membrane, as discussed above, leads to modifications of earlier estimates of the effective round trip

phase of the EM wave as it crosses the trachea and returns back to the (Burnett (1999). The simulations herein show that the high dielectric constant of the vocal fold membrane adds significant phase to the EM wave crossing the trachea. In addition, a second, smaller, phase enhancement occurs as the EM wave crosses the trachea air space, and is due to the fact that EM wave in the trachea, also extends into the surrounding high dielectric constant tissue. When these sources of additional phase accumulation are taken into account, the location of reflection becomes consistent with the glottal opening location.

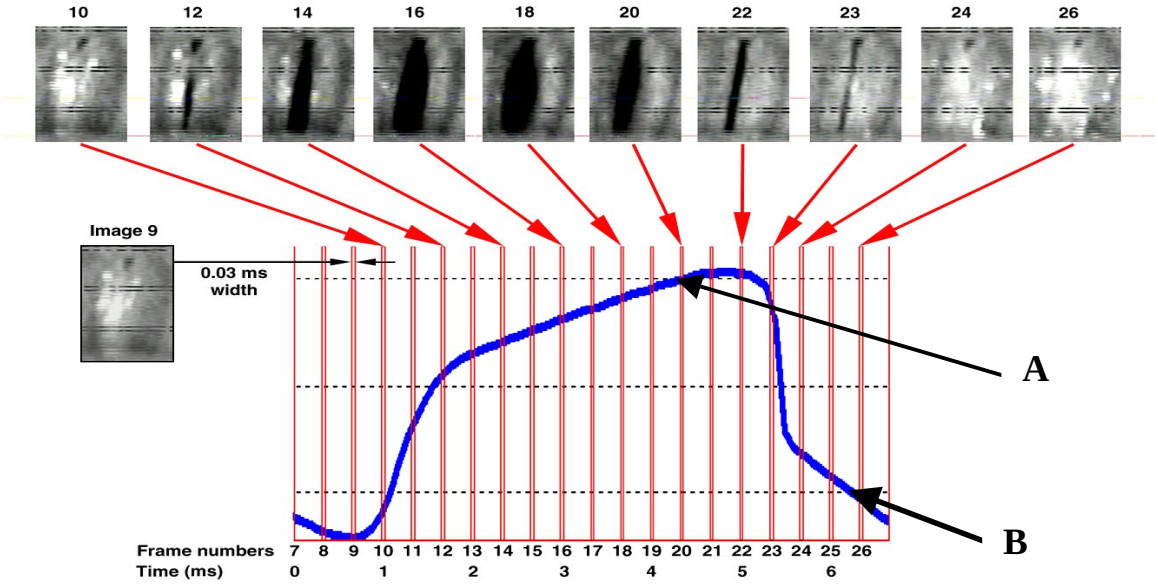

Fig. 11: Typical GEMs signal from a male subject with corresponding vocal fold images. Location A shows continued sensor signal increase, even as the inferior vocal folds are closing, and location B shows continued sensor signal after the folds have closed, indicating continued vocal fold tissue rearrangement.

It is useful to compare the superior view of the vocal fold surface tissues to the corresponding EM sensor signal (Burnett 1999, Burnett and Leonard 1999). See Fig. 11 below for an example of data using these authors' techniques. It is interesting to note that at location "A" the EM signal is still rising indicating an increase in EM wave reflectivity, while the glottis has reached its maximum open area. Increasing reflectivity is associated with continued enlargement of the glottis. Similarly, the signal at location "B" shows continued reduction in sensor values even though the corresponding images of the superior surfaces already indicate closure. This example illustrates the sensitivity of radar-like EM sensors, such as the GEMs, to many idiosyncratic factors of the subject, which may include: size, shape, dielectric constant, modes of motion, and on the time dependant arrangements of tissue layers.

The GEMs signals from vocal fold motions show individual distinctive variations as illustrated in Figs. 3 (Titze 2000). In Fig. 11 above, the structures at locations A and B, are different than those seen in Fig 4 and similar to only one of the subjects in Fig 3. These features appear to be associated with internal variations of the vocal fold structures, "sensed" by the EM wave as it is guided by the membrane. They can, in principle, be analyzed using the numerical simulation techniques employed herein. These idiosyncratic patterns may be useful for speaker verification as they provide another characteristic parameter, along with pitch and other acoustic variables, for use in speaker verification algorithms (Gable 2000).

VI. Conclusion:

Based upon the experiments and simulations described herein, and elsewhere, EM radar-like sensors (e.g., GEMs) can measure human speech articulator motions in real time. Examples are vocal folds, vocal tract walls, tongue, cheek tissue, jaw, and others. By specifically targeting either the vocal folds or the tracheal walls, these EM sensors provide information on their response to pressure and to air-flow. However the signals from these two types of tissues are very different in both shape and amplitude. EM sensor measurements taken at the laryngeal prominence location are due to the changing configurations of the vocal folds. In contrast, EM sensor measurements of air pressure

induced sub-glottal, tracheal wall movements are shown experimentally and analytically to be associated with a nominal 10-20 micrometer wall movements. This movement is associated with anterior tracheal wall motion (not posterior wall movement). Similar signals are measured from supra-glottal vocal tract walls, and from the inner walls of the oral cavity. The association of a voiced excitation function of speech (and other functions) with these signals is useful for many applications. Continued work to define the ranges of validity and accuracy is desirable.

For specific applications, the EM sensor circuitry should be optimized for the tissues that they are to measure. Some considerations are the appropriate body surface location, the area of target, the rate of movement, and the internal location of the articulator target. In addition, the presently used radar-like sensor (i.e., GEMs) is a single internal-reference interferometer that uses one antenna for transmitting and a second for receiving. While its signal vs. distance pattern has been of use for these initial measurements, the device is much too sensitive to small location movements, on the order of a millimeter or two, than is acceptable in clinical or commercial applications. Similar radar sensors are commonly made with two outputs, one called the "normal" and the second called the "quadrature" output, which is shifted 90 degrees in phase to the normal, (Skolnik 1990). The signals from these two outputs, when processed together, enable stable output versus distance to be obtained, and they provide data of the type shown above for either the trachea wall or the vocal folds (or other articulators), depending upon the user's applications.

Data from the EM sensors of the type illustrated here are usually combined with corresponding acoustic data for speech processing applications. For example, they enable robust, extremely accurate pitch detection (< 1 Hz accuracy) and acoustic signal de-noising better than 10-20 dB reduction, depending on type of noise (Aliph 2002). They also appear to enable narrow bandwidth speech compression (200-400 Hz bandwidth), and preliminary experiments indicate that speaker verification is possible with error rates less than 1:1000 (Gable 2000). In addition these sensors can be used in a non-contact mode and can measure other vocal articulator movements (Holzrichter 1999). These relatively new EM sensors are becoming increasingly economical, use

very little power, and can be built into a variety of instruments and commercial devices.

Acknowledgements

We would like to thank Dr. Greg Burnett for many early suggestions and measurements. We would also like to thank Drs. Rebbecca Leonard and Wayne Lea, Professors Richard Freeman, Neville Luhmann, John Ohala, Ingo Titze, and Jeffrey Wadsworth for their continued interest and support. We also thank Mr. E.T. Rosenbury, D. Poland, and Drs. Steven Azevedo and Todd Gable for their assistance with the sensor designs, measurements, and analyses.

The work at LLNL was supported by the US Department of Energy, DARPA, the National Science Foundation under SGER grant 972912, and at the Massachusetts Eye and Ear Infirmary by NIH/NICDC.

APPENDIX A: The homodyne sensor

Normally, one thinks of radar as transmitting an EM wave that reflects from the skin of an object, returning to the receiver with a measured time delay that is associated with the roundtrip distance to the target (e.g., range gate mode). However for this approach to work with small distance variations, very precise time information is required, << 1 sec, and the systems become rather expensive. For situations where absolute distance is not important, but where a precise measurement of a small distance variation is required, the interferometric (i.e., homodyne) radar mode is most effective. In this mode, the reflected wave is compared to a local wave using a "mixer" within the sensor, yielding an output that varies with the cosine of the relative phase of the two signals, ϕ. (Skolnik 1990). In optical interferometers similar methods yield the well known "fringe patterns" with a signal maximum and minimum every 1/4 wave of phase difference between the reflected wave and the local reference (see Fig. 10).

In these systems, the locations of the reflecting surfaces are indeterminate to 1/4 wave. When using EM waves inside tissue, account must be taken of the shortening of the wave, in inverse proportion to the higher index of refraction, which is (dielectric constant)$^{1/2}$ = $\varepsilon^{1/2}$. In muscle tissue, with a high water content, $\varepsilon = 50$, the index becomes $\varepsilon^{1/2} = 7$. When cartilage and other types of tissue with less water content are averaged, we use a dielectric constant of 25, and an index of 5. Then the nominal 13 cm wave in air shortens to about 2.6 cm in the tissues around the trachea. When tissue movement is measured at the most sensitive part of the "fringe" (i.e., using AC detection at the dark fringe location), movements < 5 micrometer are easily detected and converted into tissue movement signals for acoustic purposes. Thus, these sensors have a resolution of at least 1 in 10^4.

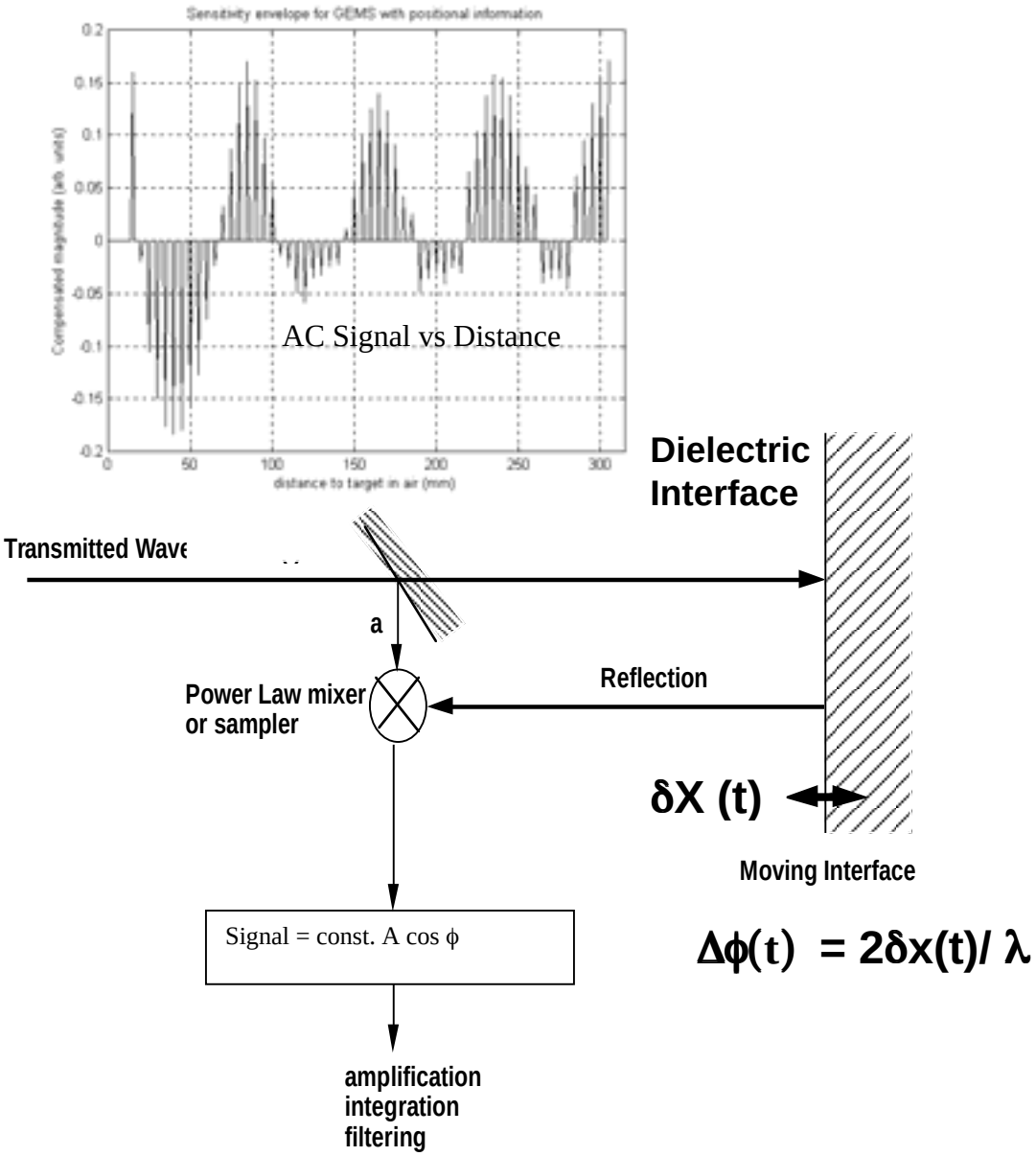

Fig. 10: Illustration of a radar interferometric measurement where the plane represents a reflection from a tissue interface. The beam splitter represents a method of obtaining a reference wave, which in an electronic circuit can be a signal tap effected by a resistor or other circuit elements. The mixer is usually a diode, in which the reference and reflected beam are multiplied leading to a "base band" (i.e., low frequency sensor output signal) and a 2nd harmonic signal (e.g., at 4.6 GHz) of the radar wave that is not used. The base band signal is proportional to the amplitude A (of the product of the amplitudes of the reference*reflected waves) times the cosine of the phase difference, ϕ, between the reference and the reflected waves. This gives the "average" reflectivity of the targeted interface versus time. The data insert is from Burnett 1999.

When the present sensor is placed near the human body, all dielectric interfaces in the beam path will reflect some of the wave energy. However, the multiple sources and the interferometric ambiguities are often not important, because only one moving tissue interface, with a specific frequency of movement, is in the field of view of the sensor's EM wave. The EM sensors used in experiments such as the ones described here are AC filtered, producing signals only from those interfaces that move at rates that pass through one or more filter bands. In the GEMs sensor, signals are high-pass filtered, > 70 Hz. Lower frequency signals, associated with vocal fold ad- or abduction, breathing, artery wall expansion and contraction, and other slow movements, are not measured. For other applications, GEMs-like sensors can be designed with one or more filter bands to measure slower or faster tissue movements.

A further consideration is that objects such as air tubes or vocal fold structures, when they are approximately the size or smaller than the EM wavelength (in the tissue) of about 2.6 cm, will cause the wave to diffract rather than follow simple ray-optics reflection or refraction trajectories. These effects are difficult to estimate without the aid of computer simulation. In addition the strength of the reflected wave from any given interface depends on several factors. They include primarily the cross sectional size of the reflecting interface and the magnitude of the dielectric discontinuity (e.g., from an air tube with $\varepsilon = 1$ to a smooth-muscle wall with $\varepsilon = 50$). In addition, the attenuation of the propagating wave by scattering and absorption of the tissues themselves, as well as diffraction and refraction effects must be followed for an accurate description of the targeted tissue responses.

Finally, when following a speech articulator's rapid movements (in the cases in this paper), the objective is to measure changes in the interface positions (or shapes), from one time to another. When the homodyne sensor is located with respect to the targeted tissue interface, the phase between the local and the wave reflected back to the sensor should be close to $\pi/2$ for maximum signal strength. For example, by moving the GEMs sensor by a few mm, this maximum can usually be found experimentally. Under this condition, the phase of the moving target interface becomes $\phi = \pi/2 + \Delta\phi(t)$, where $\Delta\phi(t)$ is a small phase proportional to two times the tissue target movement. Then the normal homodyne signal, which is proportional to the cosine between the

reflected signal and the reference, is cosine ($\pi/2 + \Delta\phi(t)$), but using a trigonometric identity this cosine becomes sin ($\Delta\phi(t)$) , which in turn is proportional to $\Delta\phi(t)$ for small motions. For large tissue movements (see Fig. 10), the full formula for the signal = const. A * cos (ϕ) must be used.

While the above discussion holds for single reference interferometric radar sensors, a second internal reference can be added which measures the "quadrature" reflection. (Skolnik 1990). The reason the second signal is "in quadrature" is that it measures the phase of the reflected signal that is $\pi/2$ out of phase with the normal signal. This is a common practice in such interferometric radar because it is usually not acceptable for the signal to drop to zero at a null position, or to change sign depending on the sensor-target distance. With such a quadrature system, the signal from targeted tissue can always be detected.

In addition, antennas for directing EM waves toward specific body locations have not been discussed. The GEMs sensor does not have a focusing antenna. However, an EM beam can be focused to a transverse dimension of approximately the EM sensor's wavelength in the tissue, or to about 4 cm in transverse extent for 2.3 GHz waves. Focal patterns can be obtained as needed by using the focusing properties of the high dielectric material in which the targeted tissue interfaces are located, and by appropriate antenna design.

APPENDIX B: EM wave propagation in throat structures.

A precise transfer of amplitude-of-motion versus signal-level-calibration for free-air targets, to the measurement and determination of an internal tissue element's movement is difficult. To precisely calibrate the GEMs signal vs. absolute motion of the posterior tracheal wall or a vocal fold structure, buried in a dielectric neck, requires more information than is reasonably obtainable. However approximate estimates of the "effective distance" from the sensor's antennas to the targeted tissue, the losses, the focusing effects, and the return path were made as follows. The measurements of the posterior trachea wall, at a location 3-5 cm below the vocal folds, is considered first. The EM wave leaves the sensor's transmit antenna, horizontally polarized (in the coronal plane), which radiates isotropically (Poland 2001) However, when the sensor is contacted to the neck tissue, a 1 - 10% coupling of the radiated energy to the neck tissue is estimated. As the EM wave crosses the skin, it loses about 50% of the signal due to Fresnel reflection from the low to high index of refraction interface. It is then focused by the cylindrical curvature of the neck. Also, the apparent geometrical distance from the antenna to the targeted tissue is shortened due to the high index of refraction of the tissue, n . This process is called brightness enhancement (Born and Wolf 1965) increases the signal by n^2 . These processes increase the EM wave intensity in the direction of the targeted posterior tracheal wall by a factor of about 50.

On its way to the posterior wall, the EM wave first contacts the anterior trachea wall. The reflection from this wall is associated with a movement of 5 - 20 microns with the air pressure changes, it will reflect EM waves strongly back to the sensor, giving a signal estimated and measured to be about 40 mV. After partially reflecting from the anterior wall, the EM wave continues to propagate (i.e., it evanesces) along the side walls of the tracheal air tube to the posterior wall. It loses a factor of about 3-5 in amplitude due to the incapacity of the wave to propagate efficiently directly across (i.e, inside) the air tube. Then the EM wave partially reflects (i.e., scatters) from the posterior wall, along a similar path back to the receiver antenna. As the waves leave the skin, the inverse of brightness enhancement occurs. For example, 2-D and 3-D simulations show that 1% of the signal entering the neck returns to the antenna from the posterior tracheal wall. They also show that typically 0.8 % variations of this 1%-reflected wave occur as the posterior wall moves 10-20 microns under pressure excursions.

In the case of signals associated with EM wave reflections from the vocal folds, propagation phenomena similar to those described above for the trachea begin to take place. Then the EM wave is carried into the tracheal tube by the effect of the high dielectric constant (i.e., high index of refraction) vocal fold membrane. There it both reflects from the glottal opening in the membrane and, at the same time, the wave forward scatters such that it continues to the rear wall of the larynx and then reflects back. The changes in location of the reflecting surfaces of the glottal opening, as the glottis enlarges and contracts, cause changes in both the amplitude and phase of the reflected EM waves which cause an AC sensor signal change of about 20-40%, depending on the glottal opening area (see Fig. 10). This leads to GEMs signals typically 30x greater than that those detected when the sensor is placed well below the vocal folds, where only tracheal wall movements are sensed.

REFERENCES:

Aliph (2000), applications of EM sensor/acoustic speech technologies. Aliphcom Incorporated, San Francisco, CA, see examples at www.aliphcom/sound (best accessed using Microsoft Explorer)

Born, M. and Wolf, E. (1965) "Principles of Optics" Pergamon Press, Oxford

Bruel and Kjaer vibration generator: exciter body type 4802 , with a general purpose head type 4817, and a model 2708 amplifier.

Burnett, G.C. and Leonard, R. (1999) "Use of Kodak Ektapro High-Speed DigitalCameras in Laryngoscopy" Phonoscope (2) 1, 33 (1999) Singular Publishing Group

Burnett, G.C.,(1999) "The Physiological Basis of Glottal Electromagnetic Micropower Sensors (GEMS) and Their Use in Defining an Excitation Function for the Human Vocal Tract" Thesis UC Davis, Jan. 15th, 1999, available on the Website mentioned in [1], and through University Microfilms, Ann Arbor, MI, document number 9925723.

Cranen, B. and Boves, L. (1985) "Pressure measurements during speech production using semicondutor miniature pressure transducers: Impact on models for speech production" , J.Acoust. Soc. Am. 77 () 1543 (1985)

Flanagan, J. L.,(1965) "*Speech Analysis, Synthesis, and Perception* ," Academic Press, Inc., New York. See pp. 40-41. Also in the 2nd edition in 1972.

Fredberg, J.J. and Hoenig, A. (1978) "Mechanical Responses of the Lungs at High Frequencies", J.Biomechanical Eng (100), 57 (May 1978)

Funt, Y.C. (1993) "Biomechanics - Mechanical Properties of Living Tissues" , p. 474, 2nd ed. (1993) Springer, NY

Gable, T.J. (2000)"Speaker Verification Using Acoustic and Glottal Electromagnetic Micro-power Sensor (GEMS) Data" Thesis, Dec. 2000.

University of California at Davis, available through ProQuest Digital Dissertations, index number 9997362

Gabriel, S., Lau, R.W., and Gabriel, C., (1996)"The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues" Phys. Mod. Biol. 41, 2271-2293 (1996) IOP Publishing Ltd., UK

Haddad,W. S., Rosenbury, E. T., Johnson, K. B., and Pearce, F. J. (1997) "*Measurements of the Dielectric Properties of Body Tissues and Fluids at Microwave Frequencies*," to be published, Lawrence Livermore National Laboratory

R. J. Hawkins, J. S. Kallman, R. W. Ziolkowski, (1993) "Computational Integrated Photonics," in Engineering Research Development and Technology, R. T. Langland, C. Minichino, UCRL 53868-92, pp. 1.7-1.11, Lawrence Livermore National Laboratory, Livermore, 1993. This contains a discussion of TSARLITE by Dr. J.S. Kallman.

Holzrichter, J.F. (1995) "New Ideas for Speech Recognition and Related Technologies", Lawrence Livermore National Laboratory Report, UCRL-UR-120310 , 1995

Holzrichter, J.F., Burnett, G.C., Ng, L.C., and Lea,W.A. (1998) "Speech Articulator Measurements Using Low Power EM Wave Sensor" *Journal Acoustic Society America* 103 (1) 622, 1998. Also see the Website http://speech.llnl.gov/

Holzrichter, J.F and Burnett, G.C. (1999) "Human Speech Articulator measurements using low power, 2 GHz Homodyne Sensors" 24th international conference on infrared and millimeter waves" Lombardo. Lynette A. Ed Monterey, CA. Sept 5 (1999), Kluwer Publishing; see also Lawrence Livermore Laboratory report UCRL-JC-134775

Ishizaka, K., Matsudaira, M., Kaneko, T., (1976) "Input acoustic-impedance measurement of the subglottal system" J. Acoust. Soc. Am., 60 (1), 190 (July 1976)

Ishizaka, K., French, J.C., and Flanagan, J.L., (1975) "Direct Determination of Vocal Tract Wall Impedance," IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-11 (4), 370 (August 1975)

McEwan, T.E., (1994) U.S. Patent No. 5,345,471 and U.S. Patent No. 5,361,070 (1994).

McEwan, T.E., (1996) U.S. Patent No. 5,573,012 (1996).

Miller, E.K., Medgyesi-Mitschang, L, and Newman, E.H. , eds (1992) "Computational Electromagnetics-Frequency-Domain Method of Moments" IEEE Press, New York, 1992

Ng, L.C. (1995) Private communication regarding the EM sensor signal as an estimate of a voiced excitation function

Ng, L. C.; Burnett, G. C.; Holzrichter, J. F.; and Gable, T. J. (2000) "De-noising of Human Speech Using Combined Acoustic and EM Sensor Signal Processing", Icassp-2000, Istanbul, Turkey, June 6, 2000

Poland, D.N. and Rosenbury, E.T. , (2001) "Power and directionality of a micro-power, pulsed homodyne EM sensor" private communication concerning sensor circuitry and transmission/reception properties of the GEMs sensor

Roark, R.J. and Young, W.C., (1975) "Formulas for Stress and Strain", 5th ed., 1975, McGraw-Hill Book Co., NY

Rosowski, J.J.; Ravicz, M.E.; Teoh, S.W.; and Flandermeyer, D., (1999) "Measurements of Middle-Ear Function in the Mongolian Gerbil, a Specialized Mammalian Ear", Audiol Neurootol 1999;4:129-136(1999)

Rothenberg, M. (1981) "Some relations between glottal flow and vocal fold contact area," ASHA Rep. 11, 88-96 .

Sharpe, R.M., J.B. Grant, N.J. Champagne, W.A. Johnson, R.E. Jorgenson, D.R. Wilton, W.J. Brown, and J.W. Rockway, (1997) "EIGER: Electromagnetic Interactions GeneRalized" IEEE AP-S

International Symposium and North American URSI Radio Science Meeting, Montreal, Canada, July 1997, pp. 2366--2369.

Skolnik, M. (1990). "*Radar Handbook*," 2nd edition., McGraw-Hill, New York

Stevens, K.N. (2000) "Acoustic Phonetics" MIT Press, Cambridge, MA, 2000

Svirsky, M.A., Stevens, K.N., Matthies, M.L., Manzella, J. Perkell, J.S., and Wilhelms-Tricarico, R. (1997) "Tongue surface displacement during bilabial stops", J.Acoust. Soc. Am. 102 (1), 562 (1997)

Titze, I. R.(1984). "Parameterization of the glottal area, glottal flow, and vocal fold area" J. Acoust. Soc. Am. 74(2), 570-580

Titze, I.R., (1990) "Interpretation of the Electroglottographic Signal" , J. of Voice 4 (1), 1-9 (1990) Raven Press, Ltd. NY

Titze, I.R., (1994)"Principles of Voice Production" Prentice Hall, NJ, 1994

Titze, I.R., Story, B.H., Burnett, G.C., Holzrichter, J.F., Ng, L.C., Lea, W.A., (2000) "Comparison between electroglottography and electromagnetic glottography" 107 (1), 581 (January 2000)

Whirley, R.E. (2001), private communication on estimating Young's modulus for smooth muscle from biomechanics data sets.